Machine Learning for Natural Language Processing

Arieda Muço

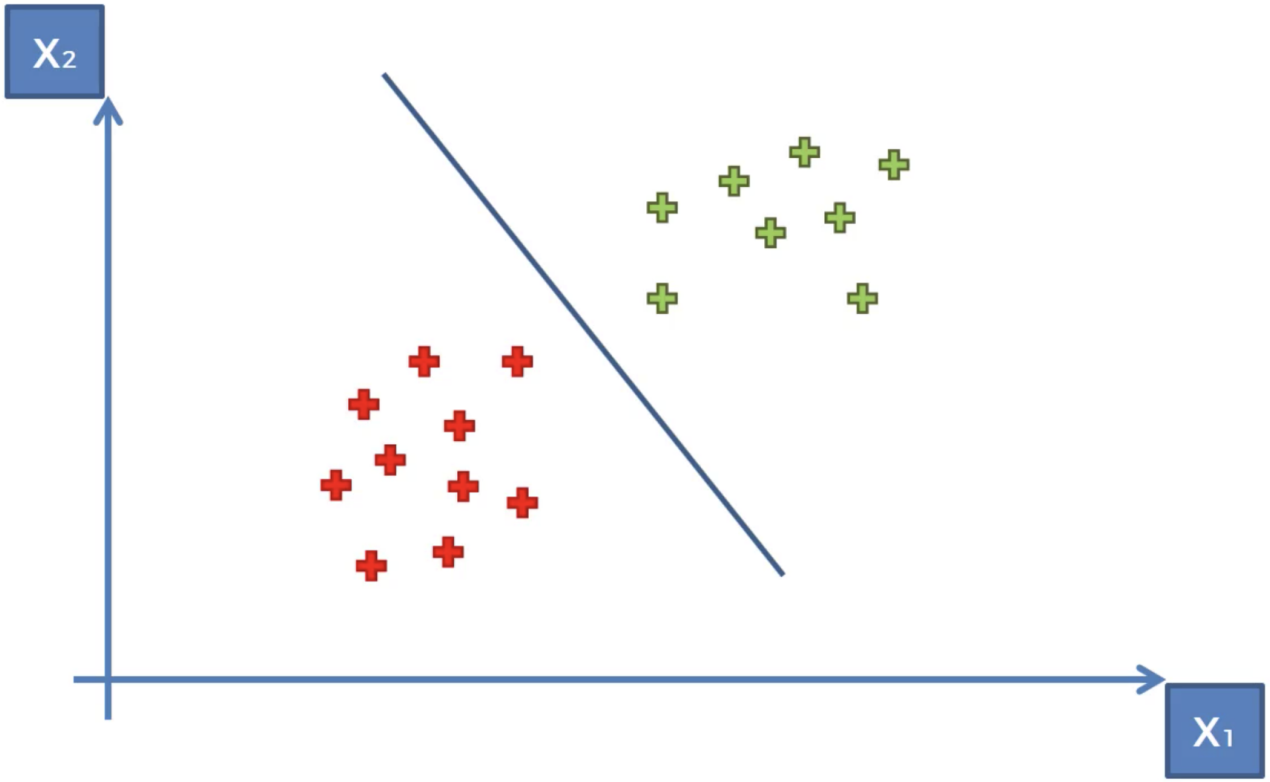
Central European University

March 18, 2022

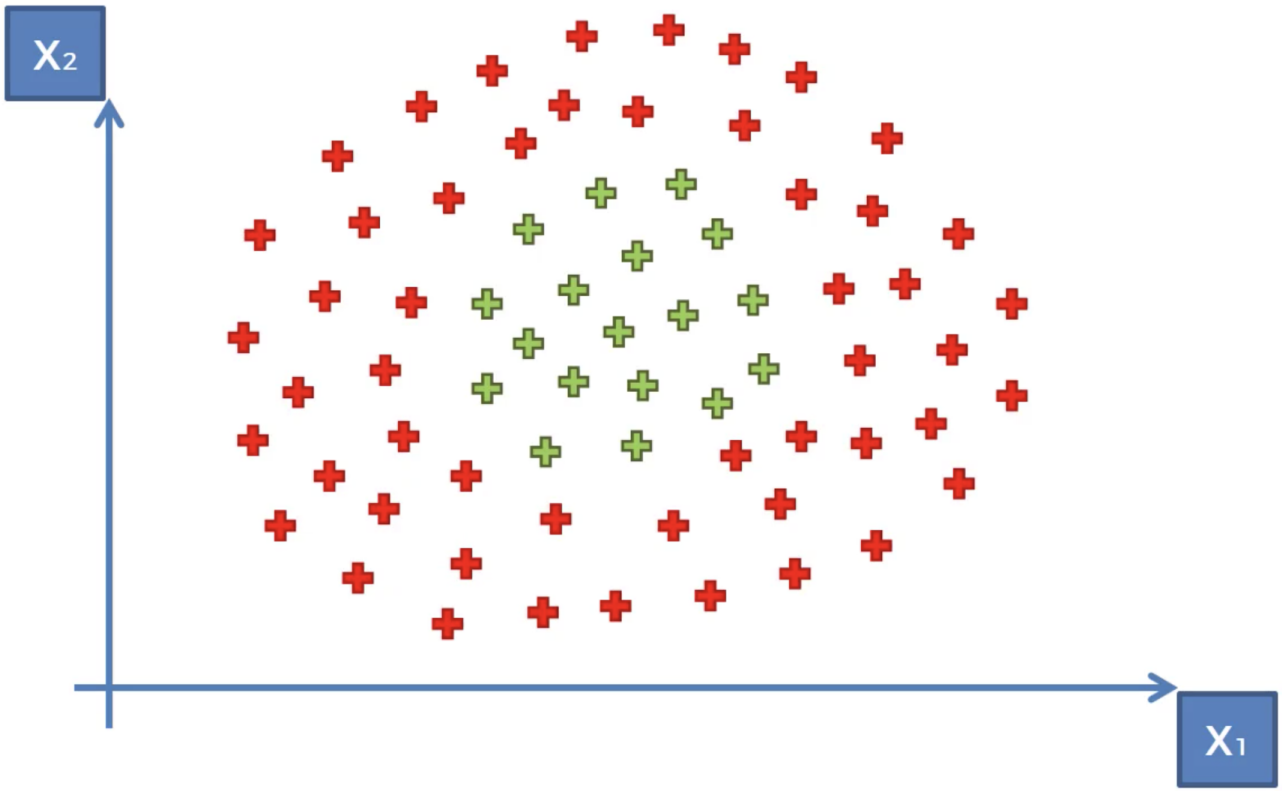
Support Vector Machines

- Support Vector Classifier is an supervised learning algorithm that allows to separate observations that fall in different classes.
- How?
  - Mapping into a higher dimensional space
  - Kernel Trick

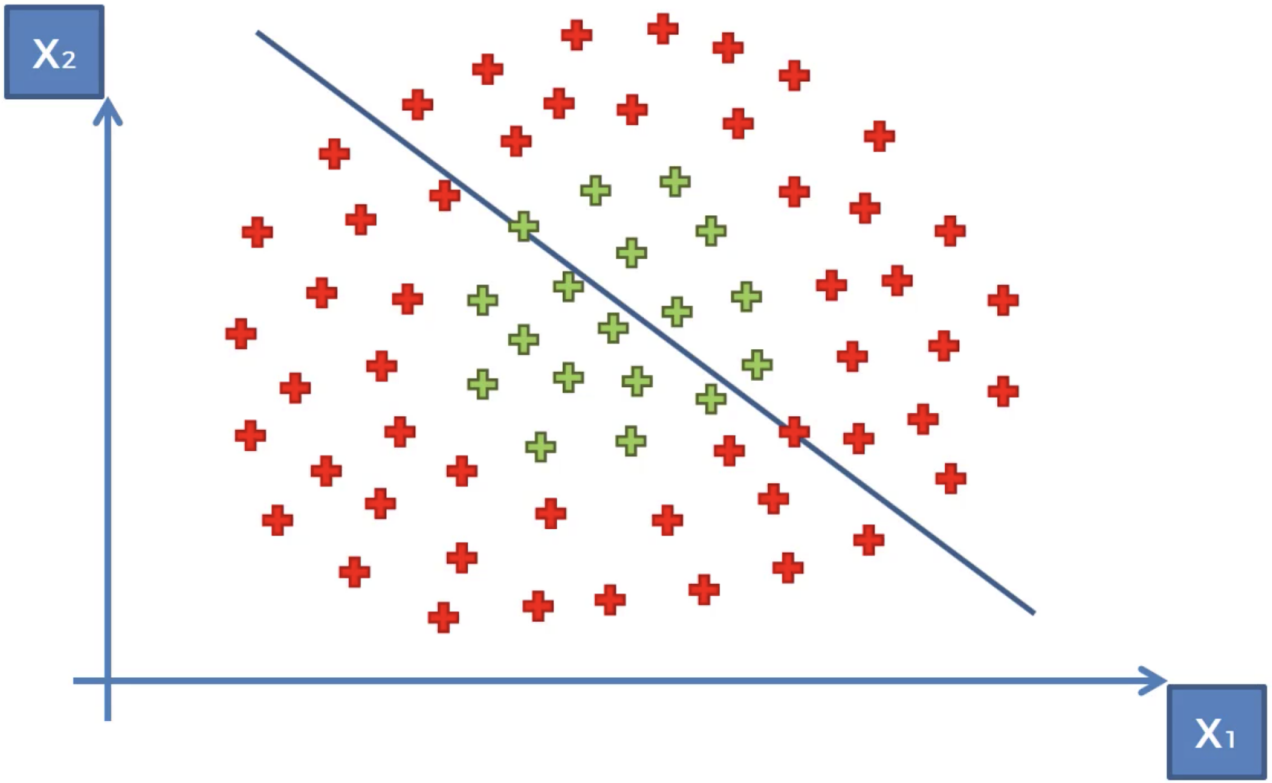
Linearly Separable

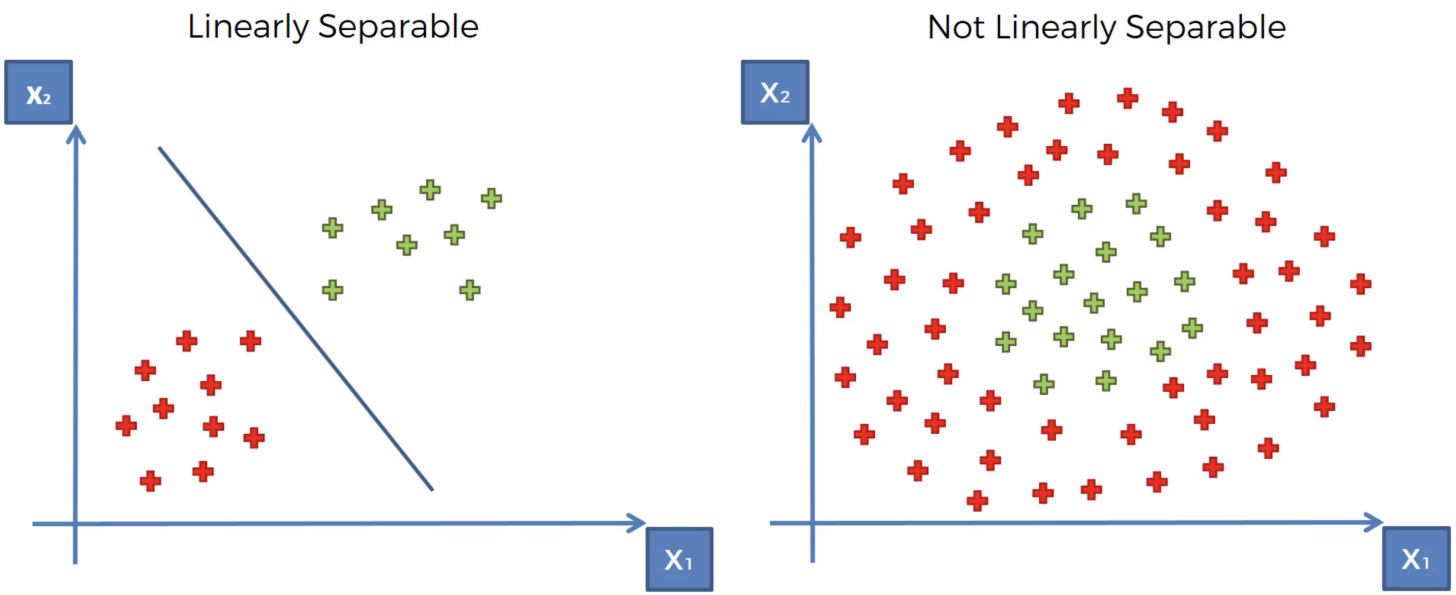


Non Linearly Separable

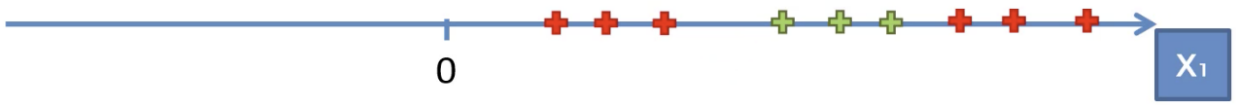


Non Linearly Separable

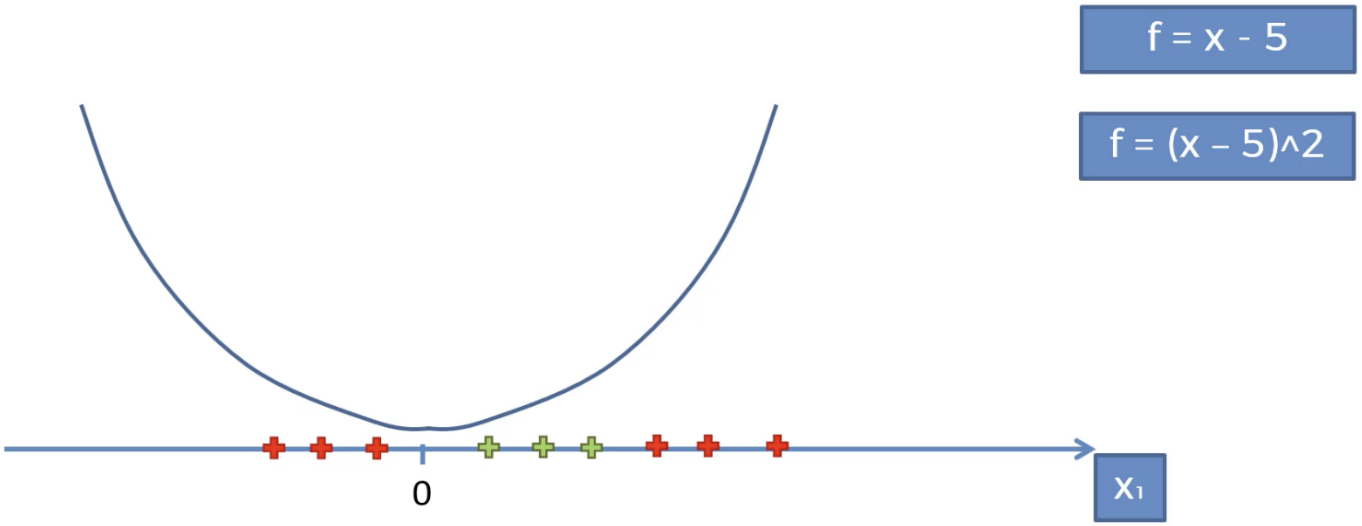



Linearly Separable
X2
X1
Not Linearly Separable
X2
X1

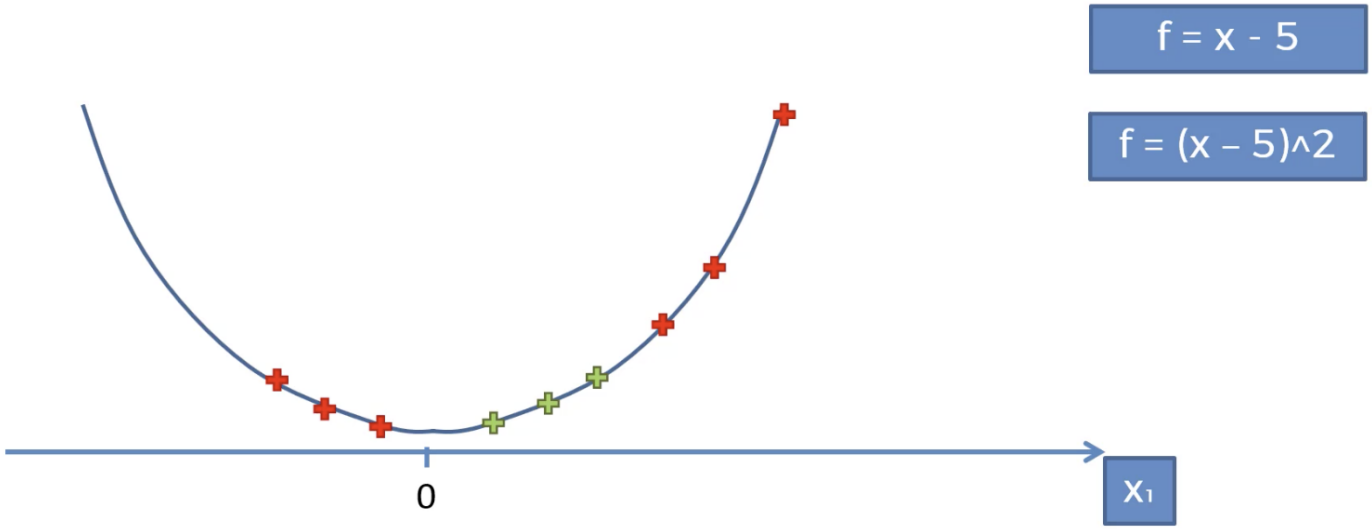
Mapping to a higher dimension



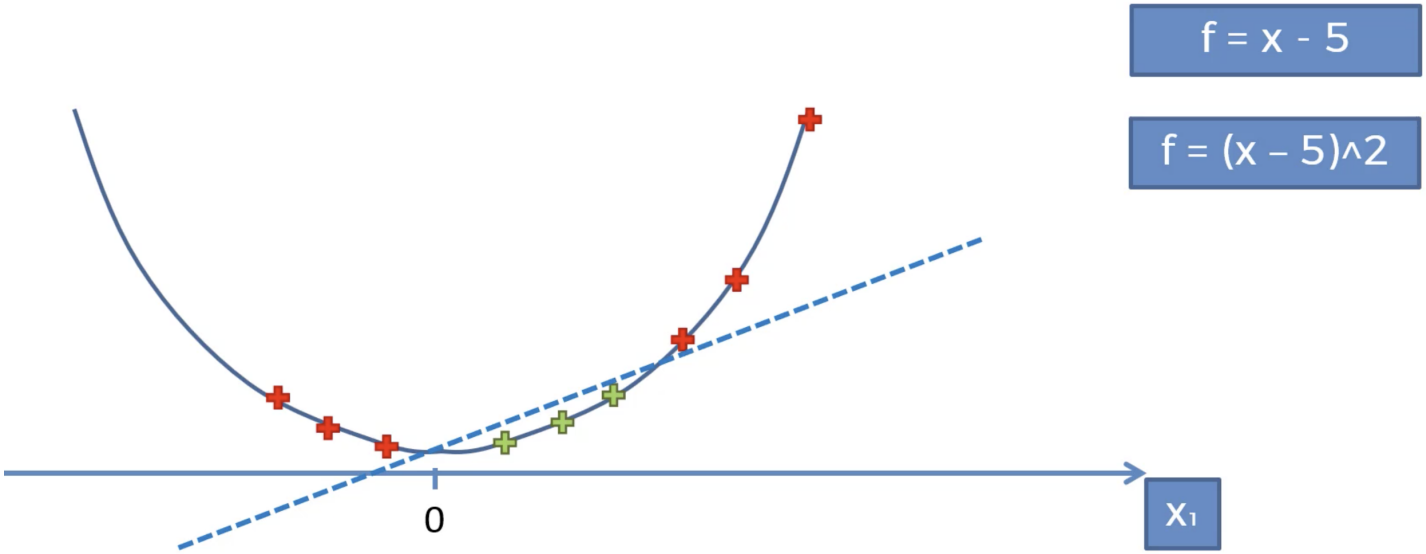
Mapping to a higher dimension



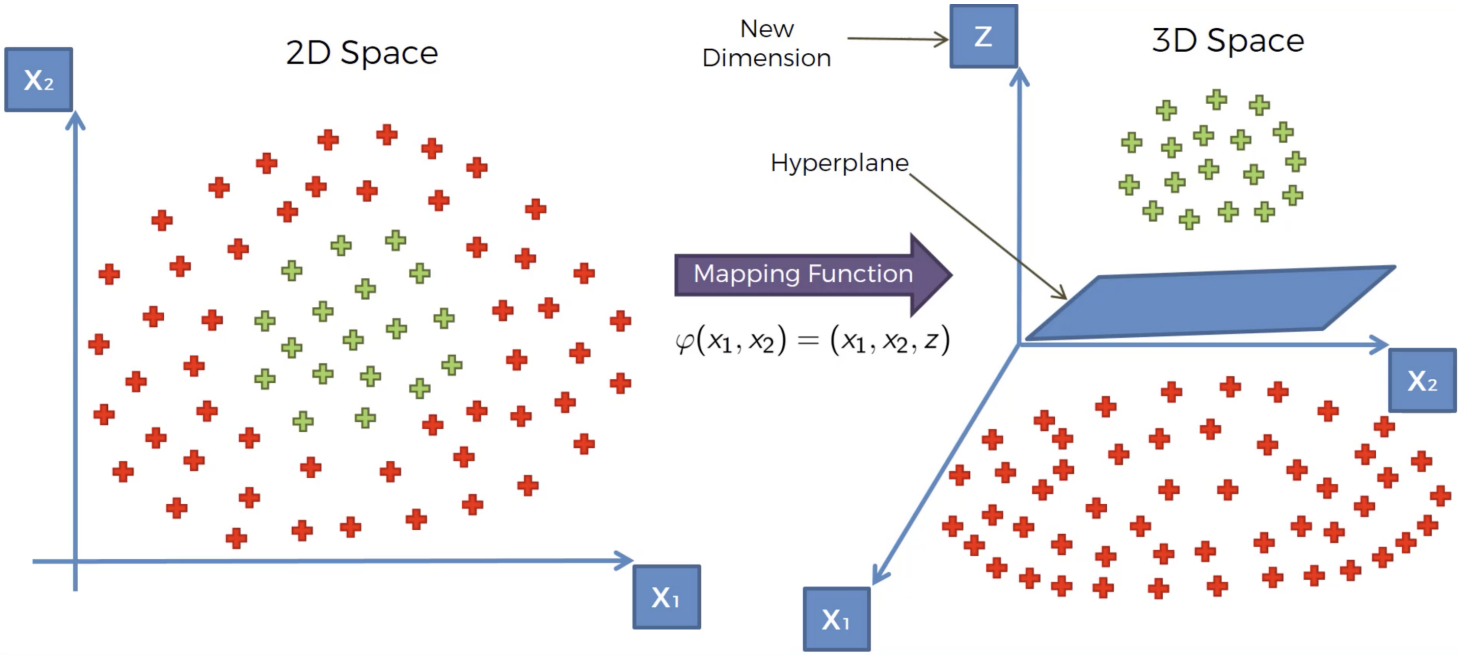
Mapping to a higher dimension



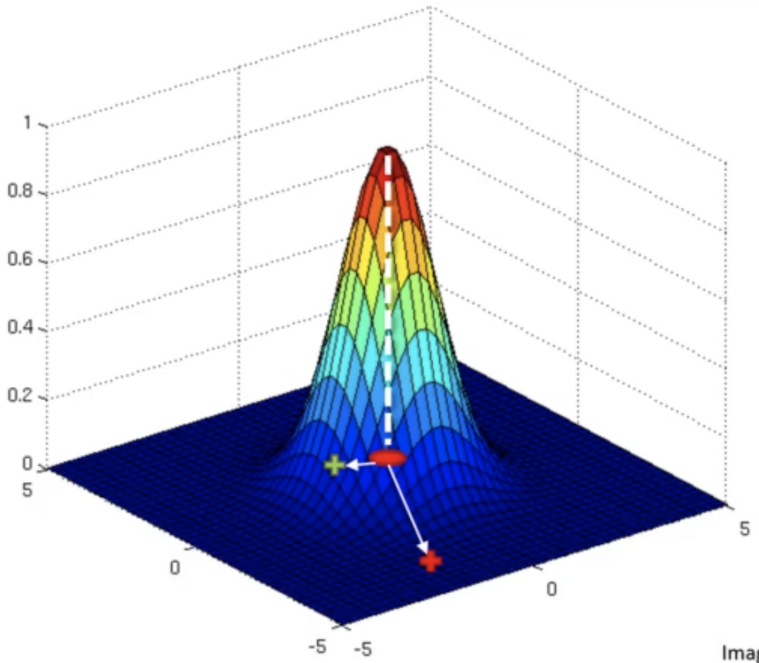
Mapping to a higher dimension



Mapping to a higher dimension



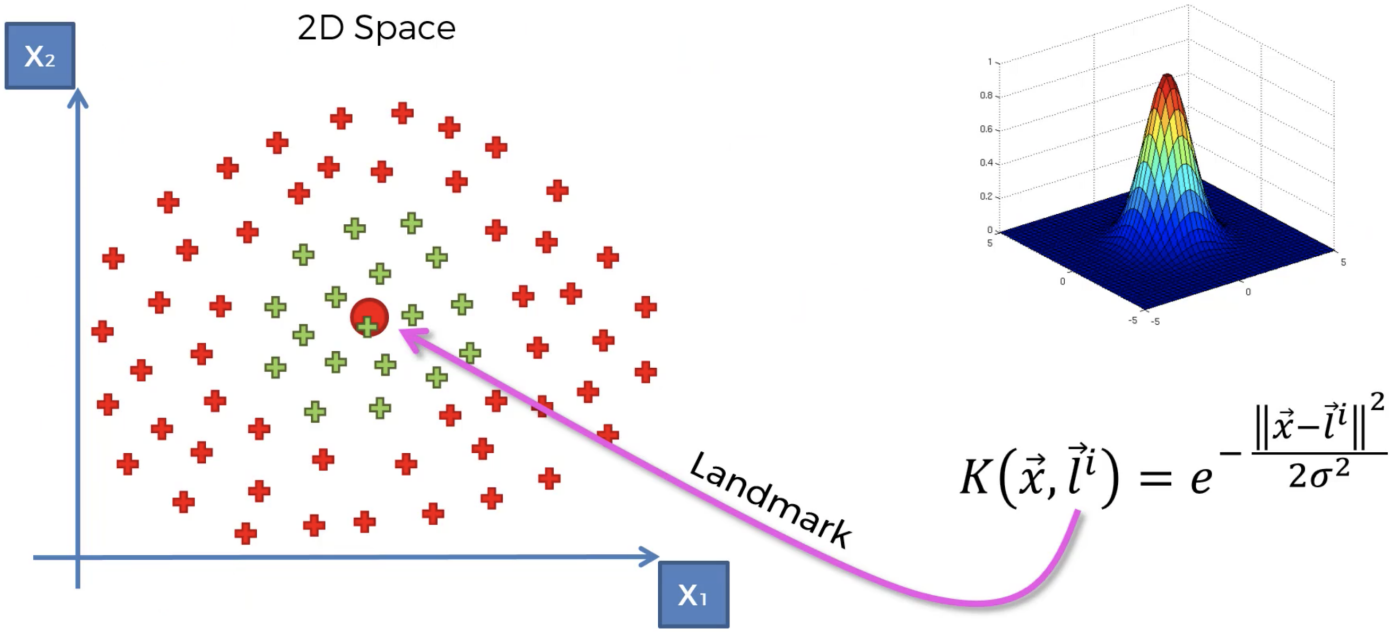
The Kernel Trick: RBF-Kernel



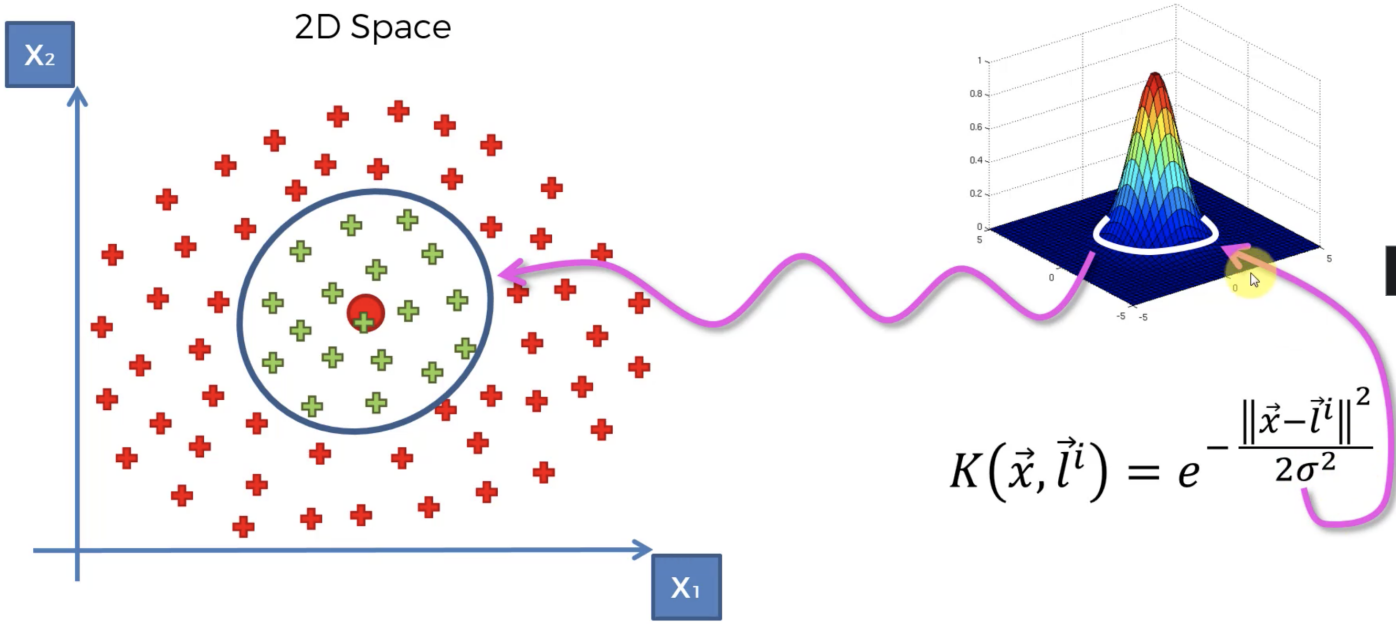
$$K(\vec{x}, \vec{l}^i) = e^{-\frac{\|\vec{x}-\vec{l}^i\|^2}{2\sigma^2}}$$

Image source: http://www.cs.toronto.edu/~duvenaud/cookbook/index.html

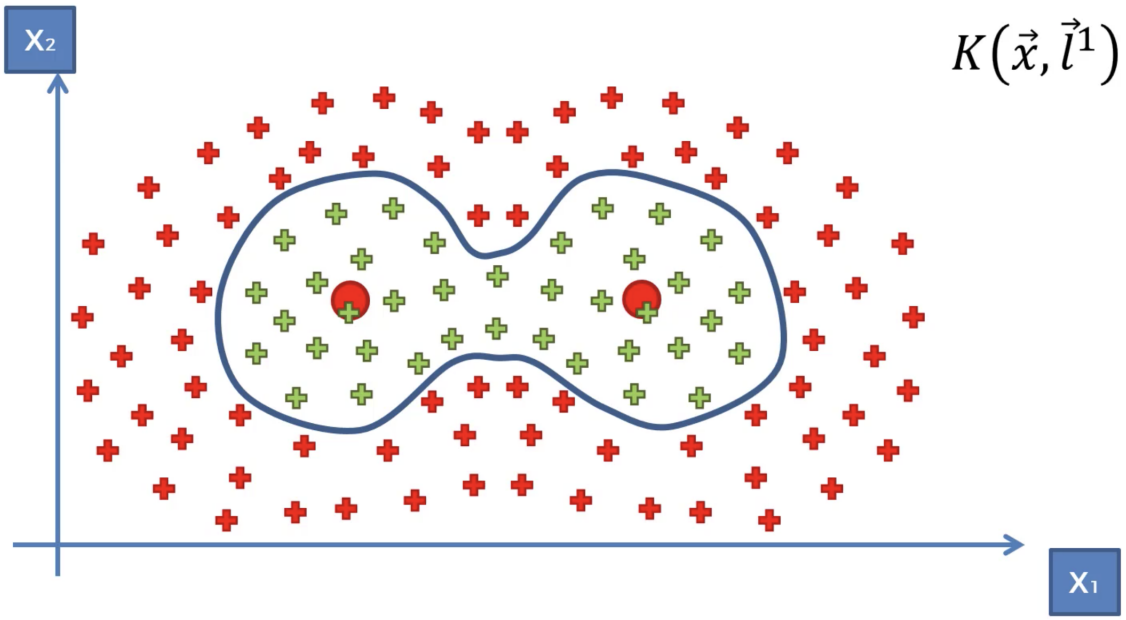
RBF-Kernel


2D Space
X2
X1
Landmark
$K(\vec{x}, \vec{l}^i) = e^{-\frac{\|\vec{x}-\vec{l}^i\|^2}{2\sigma^2}}$

RBF-Kernel



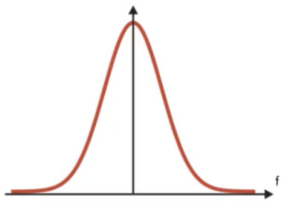
RBF-Kernel



$$K(\vec{x}, \vec{l}^1) + K(\vec{x}, \vec{l}^2)$$

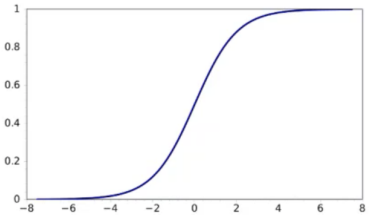
(Simplified Formula)

Most Common Kernels



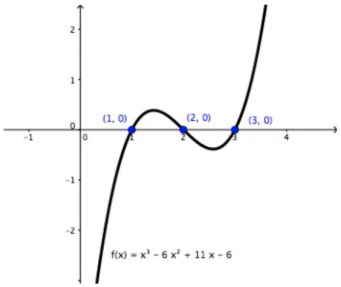
Gaussian RBF Kernel

$$K(\vec{x}, \vec{l}^i) = e^{-\frac{\|\vec{x}-\vec{l}^i\|^2}{2\sigma^2}}$$



Sigmoid Kernel

$$K(X, Y) = \tanh(\gamma \cdot X^T Y + r)$$



Polynomial Kernel

$$K(X, Y) = (\gamma \cdot X^T Y + r)^d, \gamma > 0$$